PHNL000081
US

NL 000139
I

Eur päisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

00200540.3

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 08/11/00
LA HAYE, LE

Europäisches Patentamt
Eur pean Patent Office
Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 00200540.3
Demande n°:

Anmeldetag:
Date of filing: 17/02/00
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
A method of making highly-localized lateral channel profile of Si MOSFET's coupled to gate workfunction engineering

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

H-NL000081 EP-P 17.02.2000

A meth d of making highly-localized lateral channel pr file of Si MOSFET's coupled to gate workfunction engineering

Y. V. Ponomarev, M.N. Webster and C.J.J Dachs

It has been shown recently that so-called "Replacement-gate" processing [1] can be highly advantageous for future generations of Si CMOS technologies. The method allows easy integration of high-k dielectrics (to elevate the problem of SiO_2 scaling limits) and metal gates (to avoid device degradation caused by "poly-depletion" in both N- and P-MOSFET's and boron penetration in PMOSFET's).
The use of metal gates can deliver significant improvement to the MOS device performance if coupled to an appropriate for the gate workfunction channel profile: most interesting metals for CMOS integration posses electron/hole workfunctions closer to the Si midgap, which allows reduce the channel doping levels for the same MOSFET threshold voltage [2,3]. Unfortunately, the reduction in the channel dope gives rise to acute "Short-channel effects"(SCE).
It has been shown [4] that highly non-uniform lateral profiling (strong and very abrupt "halo's", or "p ckets", or "envelopes", around source/drain extensions and very low doping in the channel – see figure 1) coupled to the gate workfunction change (poly-SiGe for PMOSFET's) is extremely effective in combating the SCE. Abruptness of the "envelopes" is the key in combining these profiles with advantages of low channel doping in the middle of the device (otherwise for smaller transistors the channel doping will be determined by the tails of envelopes [2]).

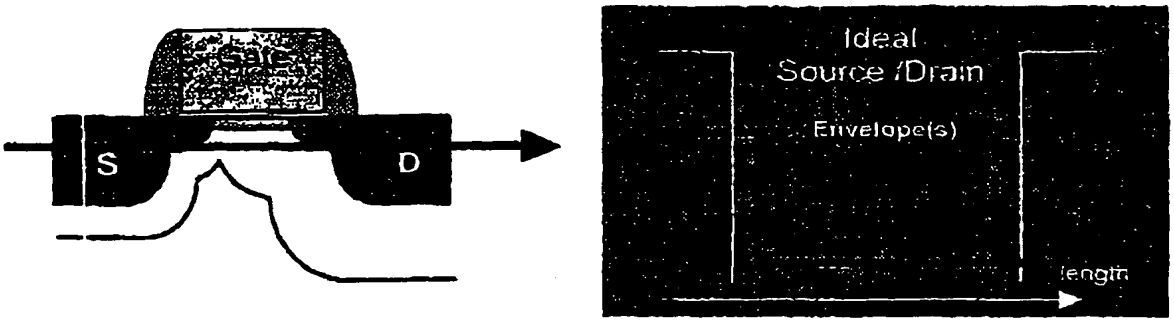


Figure 1. Short-channel effects (SCE) are caused by increased drain influence on the depletion regions of the channel and eventually source side (left-had side of the figure). The lower the channel doping, the higher the influence. Highly-doped abrupt "envelopes" around source and drain (right-hand side lateral profiles cross-section) diminish the SCE and allow the usage of low channel doping in between them.

For NMOSFET's, boron is normally used as a channel dopant, and since this impurity is prone to strong diffusion in Si, the condition of abruptness is very difficult to satisfy, since normally the "envelope" implant is performed at the same time as source/drain extensions are implanted, implying that a significant amount of consecutive processing steps will smear out the profiles. As an example, the lateral channel profiles c mputed for 0.13μm NMOSFET are presented in figure 2. Already at these gate length the channel pr file is completely determined by the lateral diffusion of b ron "envelopes" making it impossible to use gate workfunctions which differ fr m

n-type poly by more than 0.3eV (mid-gap is at ~0.56eV for Si, and will require much lower doping). This problem obviously becomes even more severe for smaller gate lengths.

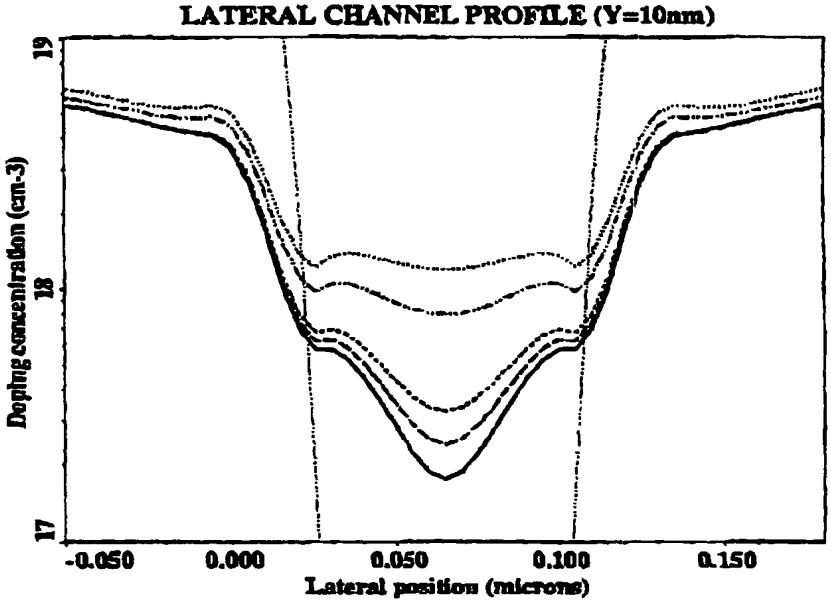


Figure 2. Boron envelopes lateral profiles for different processing conditions. The best variant achievable with "standard" processing only allows for the use of 0.3eV shift in gate workfunction.

Another disadvantage of standard approach is that the whole source/drain regions are implanted with "envelope" dopants resulting in source/drain partial compensation and a significant increase in the parasitic series resistance.

We show below an approach which can resolve these issues in a simple manner.

The process starts with low-doped substrate (1e15cm^{-3}) to ensure appropriate long-channel V_T when close to mid-gap gate workfunction material is used (TiN for NMOS). Then a complete "Replacement Gate" processing is done (see for details [6]) without any channel profiling. As a result we have an essentially complete MOSFET structure without a gate electrode.

The next step is to deposit the following stack: poly-SiGe/poly-Si (~30nm/170nm; exact thickness ratio can be optimized for a given process generation). Then these layers are "CMPed" back to the level of the surrounding oxide (figure 3(a)).

The next step is to selectively remove the poly-SiGe side-walls in the gate to a certain depth. This can be done using either an anysotropic plasma etch (etch rate is much higher for poly-SiGe than for poly-Si) or a highly selective wet etch [5]. As a result an implantation mask for the "envelopes" is created (figure 3(b)). Now, for NMOSFET's, boron can be implanted (at 0°tilt to allow extra channeling in vertical direction) and briefly annealed to result in extremely sharp envelopes around the tips of the source/drain regions (see figure 4). Such an implantation is also not expected to harm the gate oxide [7].

The back-end processing then follows: first, selective to the oxide removal of both poly-Si and poly-SiGe layers [5]. Then deposition of TiN gate electrode and, finally, contact and interconnect formation.

The suggested processing can be elaborated by the following steps:

- It is also possible to deposit first a padding layer of poly-Si before the poly-SiGe layer to get an extra freedom in lateral position of the envelope implant.

- TiN electrode use is quite problematic for PMOSFET's (gate workfunction is too low, resulting in formation of highly unfavorable "buried"-channel devices), so one can deposit poly-SiGe in-situ doped with boron, and then rem ve it only from NMOSFET's regions (one extra masking step will be required) and deposit TiN on top of that. This will result in TiN gate for NMOSFET's and poly-SiGe gate for PMOSFET's

[1] A. Chatterjee *et al.* "Sub-100nmgate length metal gate NMOS transistors fabricated by a replacement gate process" *Proc.* IEDM'97, p.821
[2] C.J.J. Dachs *et al* "Gate workfunction engineering for deep-sub-micron CMOS", *Proc.* ESSDERC'99, sept.1999.
[3] M.N. Webster *et al* "Metal Gates for 0.15 μm CMOS and beyond" *Proc.* ESSDERC'99, sept.1999.
[4] Y. V. Ponomarev *et al* "An efficient lateral channel profiling of poly-SiGe-gated PMOSFET's for 0.1μm CMOS low-voltage applications" *Proc.* VLSI Symposium, p.65 (1999).
[5] Y. V. Ponomarev "A method to vary a threshold voltage of MOS transistor by selective incorporation of SiGe in the gate electrode", *Invention Communication 25.068*
[6] P.A. Stolk, Y. V. Ponomarev "Fabrication of MOS transistors with laterally graded gate workfunction in a replacement gate process", EP Application No. 99201869.7
[7] Y. V. Ponomarev *et al.* "Channel Profile Engineering of 0.1μm-Si MOSFETs by Through-the-Gate Implantation", *Proc.* IEDM'98, p.635

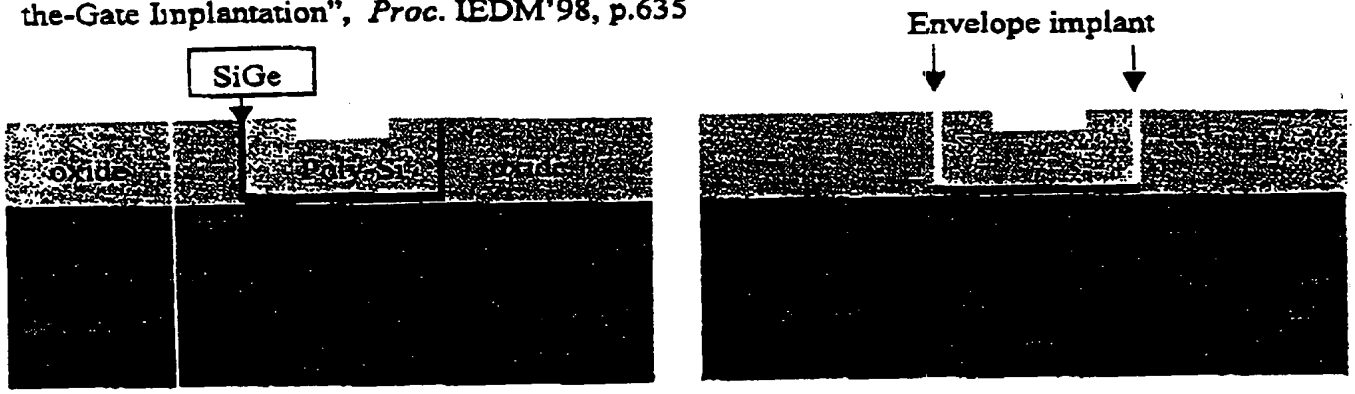


Figure 3. Boron envelopes lateral profiles for different processing conditions. The best variant achievable with "standard" processing only allows for the use of 0.3eV shift in gate workfunction.

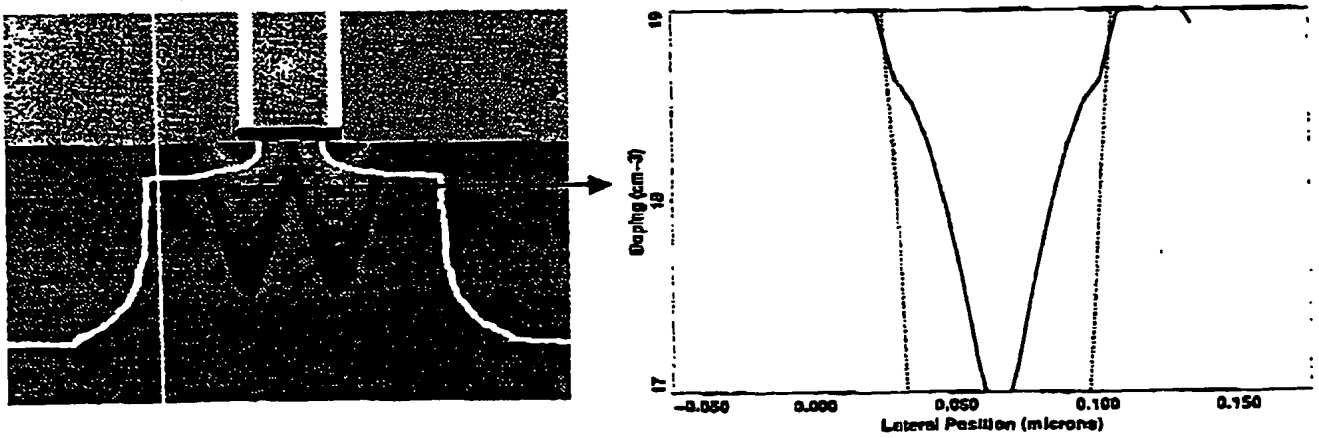


Figure 4. Results of the simulated boron profiles. By comparing the right-hand side plot with the figure 2 above, one can see that the lateral steepness of the boron profile is significantly improved.

PH-NL000081 EP-P 17.02.2000

A method to vary a threshold voltage of MOS transistors by selective incorporation of SiGe in the gate electrodes

Youri V. Ponomarev

Recently, there has been quite a significant interest to the use of poly-SiGe as a gate material for advanced CMOS transistors (see e.g. [1] and references therein). Process compatibility with existing Si technology has been shown. The change in the Ge mole fraction controls the change in the gate-semiconductor workfunction difference which, coupled to a channel profile tuning, results in a significant improvement of the PMOST performance. This performance growth is in general proportional to the Ge mole fraction in the gate and e.g for 0.18μm generation reaches 35% for 55% Ge [1].

For NMOS transistors use of poly-SiGe gates does not influence the performance unless a mole fraction of Ge higher than ~40% is used. In this case, the performance of the N-type gate becomes a problem: arsenic and phosphorus ions, conventionally implanted and diffused in the gates to dope them, become increasingly electrically inactive for increased Ge mole fraction. For instance, for 50% Ge in the gate, only 30% of dopants will be activated resulting in strong gate depletion. The situation only worsens when Ge percentage is increased to the attractive for PMOSTs 60-80%. This factor has led some authors to suggest the use of a "single-flavor" SiGe gate doping: *p*-type (boron activation is improved by an increase in the Ge percentage) [2,3]. In this case, however, the NMOS becomes a so-called "buried channel transistor". It is well-known that this type of devices are not suitable for advanced CMOS applications: weaker gate control means that the drain has a strong influence on the device parameters resulting in poor short-channel behavior, high subthreshold slope and high off-state leakage currents [4]. This approach is not satisfactory: it sacrifices NMOST performance in favor of PMOST.

Another interesting issue that has risen recently concerns the use of so-called "double-V_T " transistors in modern ICs. A possibility to use high-V_T devices (meaning, also, much lower stand-by leakage currents) in the parts of the circuits where low-power consumption is vital, and low-V_T devices (high stand-by power consumption) in the high-performance parts of the circuit can save a significant amount of power consumed by the total IC (~250 times for an average processor as cited in [5]). This could be quite an advantage, especially if the battery life is of concern. For PMOS devices "off"-state leakage currents are even more important than for NMOSTs since this type of transistors has to be designed with bigger on-chip size to compensate for a weaker current drive in the "on"-state. Usually, the double-V_T approach is implemented by using an extra mask and V_T correction implantation in the early stages of processing.

To elevate the problem of incompatibility of SiGe gates (Ge% higher than ~40%) with "double-flavor" poly-gate CMOS process, and to implement an alternative way of making double-V_T PMOS transistors at the late stages of

processing, a new sequence of the gate formation is suggested here which utilizes the following technological properties of SiGe layers: they can be removed by wet etch with high selectivity to Si [6] and that the etch-rate during plasma-etch of the poly-gates increases with Ge% [7].

Figure 1 below shows schematically the new approach implemented in the CMOS gate-stack flow to manufacture NMOS transistors with poly-Si gates (where excellent As and P activation is possible) and PMOS transistors with essentially poly-SiGe gates (SiGe layer at the bottom of the gate controls the gate workfunction and thus V_T) with high Ge%.

The process description starts after active regions of the devices are formed (gate oxide is grown too). Then the standard poly-SiGe deposition scheme for high Ge% should be used: a thin (less than 2nm) amorphous Si layer is grown first (serves as an "adhesive" layer to decrease the nucleation time and decrease roughness of the grown SiGe layer), then ~20-30nm of poly-SiGe is grown. SiGe can be grown in-situ doped with boron (since it is going to stay only on PMOS areas) to ensure excellent gate activation for PMOSTs.

The standard SN mask (already exists and is used to protect PMOS devices from NMOST source/drain implantations) is then applied and SiGe layer is selectively removed from NMOS regions leaving the thin a-Si layer on top of gate oxide intact (as well as the gate oxide itself). It is possible in the HNO_3:H_2O:dHF(0.08%HF) 35:20:10 proportion solution, where selectivity of better than 100/1 have been reported [6] for 30% Ge.

Then, a ~100nm blanket poly-Si layer is deposited to form a complete gate stack for both device types. The gates patterning is done in one go with a standard plasma etch used to pattern poly-Si layers: this can be accomplished owing to an increased speed of the etching process when Ge% is increased and thus this very thin layer of SiGe can be etched e.g. during standard overetch process. The end-point detection of the main etch (when the poly-Si layer on NMOS regions is etched) is even simplified, since the signal change at the beginning of the SiGe etch is in the same direction as that of SiO_2.

As a result we have a close to standard poly-Si gates for NMOSTs and SiGe gates for PMOSTs.

At the expense of an extra mask, some PMOS transistors can also be made without SiGe present in the gate: the flow is exactly the same apart from the mask used to protect selected PMOS devices from wet SiGe etch (step 2 in figure 1). The resulting PMOS transistors without SiGe in their gates will have threshold voltage lowered in absolute value by the workfunction shift caused by the Ge incorporation for PMOSTs with SiGe gates. This way of making a double-V_T transistors is advantageous over the standard where it is done by V_T correction implantations in the early stages of processing, since by changing the Ge% one can obtain any needed V_T value at the point when the gate electrode is formed. Use of gate workfunction control over V_T also results in reduction of V_T fluctuations caused by reduced number of dopants in the channel region of the devices.

The main concerns are:

- (1) the gate oxide quality of the NMOS transistors;
- (2) presence of a thin native oxide layer in between the bottom part of the gate and the top deposited poly-Si layer;

- (3) the workfunction change is determined by the Ge and Si co-diffusions)

All these points will have to be checked in practice (being done at th moment). It is, however, not expected that these are going to seriously endanger the suggested approach. The following arguments can be given:

(1) the a-Si layer must be closed on top of the oxide to be a good "seeding" base for consecutive poly-SiGe deposition, which means that it will be thick enough to protect the gate oxide from the wet SiGe etch attacking it.

(2) The native oxide layer underneath the top poly-Si layer might not be a problem at all since the diffusion of the high amount of dopants (As or B) should break through this layer quite easily (e.g. As emitter diffusion out of poly-Si through the native oxide, or boron penetration for thin gate oxide MOSFETs). It will be checked and if it does pose a problem, then the top poly-Si layer can be deposited straight after short dip into a weak HF solution or in a single wafer reactor (e.g. ASM Epsilon-1) where a hydrogen high-temperature bake can be used to remove all the native oxide before the poly-Si growth.

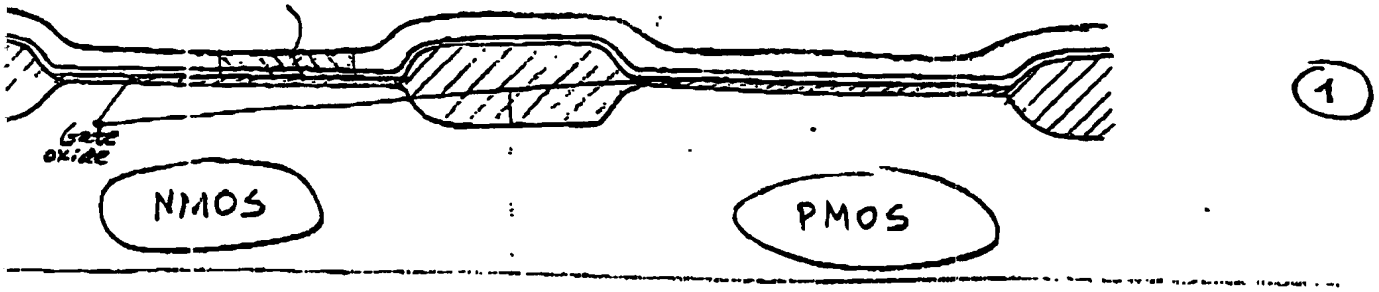
(3) The co-diffusion of Ge and Si in this triple-layer structure will result in reduction of original Ge mole fraction at the gate oxide/gate interface responsible for the Φ_{MS} value. It has been shown, however, that this process can be kept well under control for a given thermal budget endured by the gate [1].

A new approach to form gates of Si MOSFETs is suggested to provide poly-SiGe gates for only *p*-type devices where a necessary workfunction shift is needed. For *n*-type devices (and selected *p*-type devices at the cost of an extra mask at the later stages of the processing) the gates are standard poly-Si ones.

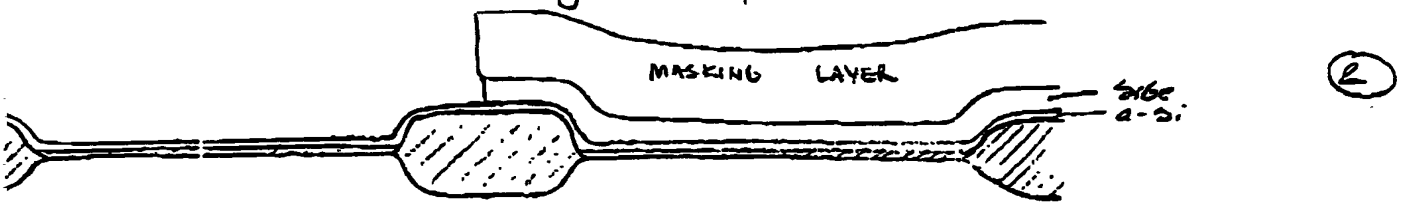
[1] Y.V. Ponomarev *et al.* IEDM'97 Techn.Digest, p.829.
[2] T.Skotnicki *et al.* ESSDERC'97 Proc., p.216.
[3] V. Z-Q Li *et al.* IEDM'97 Techn.Digest, p.833.
[4] G. J. Hu *et al.* IEEE Trans. El. Dev., ED-32, p.584 (1985).
[5] U. Ko *et al.* Symp.VLSI-TSA'96 Proc., p.275.
[6] D. J. Godbey *et al.* J. Electrochem. Soc., 139, p.2944 (1992).
[7] S. Vallon *et al.* J. Vac. Sci. Techn, 15, p.1874 (1997).

Example : CMOS : poly-Si Gate NMOST
poly-SiGe Gate PMOST

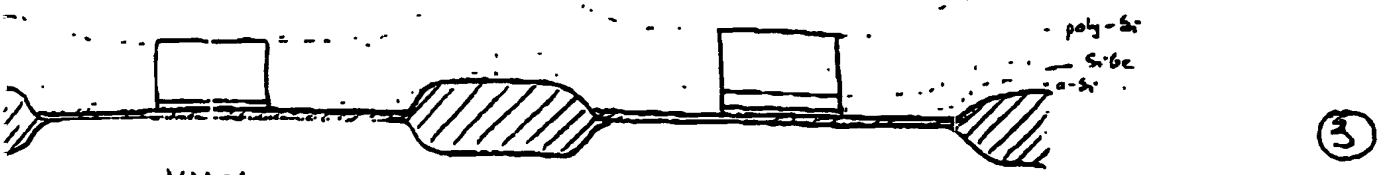
(1) deposit on wafers
a-Si/SiGe sandwich (~2nm a-Si ; ~20nm SiGe)



(2) mask the PMOS areas (with already available SN mask) and wetetch SiGe layer on the NMOS regions :



(3) deposit poly-Si on wafers (after removal of the masking layer) and use standard poly-Si plasma etch to pattern NMOS and PMOS gates



NMOS
gate stack
a-Si (~2nm)
poly-Si (100-150 nm)

PMOS
gate stack
a-Si (~2nm)
poly-SiGe (~20 nm)
poly-Si (100-150 nm)

Figure 1

PH-NL000081 EP-P 17.02.2000

CLAIMS:

1. A method of manufacturing a semiconductor device comprising a semiconductor body of a first conductivity type which is provided at a surface with a transistor having a gate insulated from a channel provided at the surface of the semiconductor body by a gate dielectric, wherein a structure is provided comprising a dielectric layer having a recess, which recess is aligned to a source zone and a drain zone of a second conductivity type provided at the surface of the semiconductor body and has side walls extending substantially perpendicular to the surface of the semiconductor body, in which recess a double-layer is applied consisting of a second sub-layer on top of a first sub-layer, which second sub-layer is etched back and which first sub-layer is selectively etched with respect to the second sub-layer and to the side walls of the recess to a depth, thereby forming trenches extending substantially perpendicular to the surface of the semiconductor body, via which trenches impurities of the first conductivity type are introduced into the semiconductor body, thereby forming pocket implants, after which at least the second sub-layer is removed, and a conductive layer is applied providing the gate of the transistor.

2. A method as claimed in claim 1, wherein, in order to provide the structure, a patterned layer is applied at the area of the planned gate, and the source zone and the drain zone of the second conductivity type are formed in the semiconductor body while using the patterned layer as a mask, after which the dielectric layer is provided in a thickness which is sufficiently large to cover the patterned layer, which dielectric layer is removed over part of its thickness by means of a material removing treatment until the patterned layer is exposed, after which the patterned layer is removed, and the recess is provided aligned to the source zone and the drain zone.

3. A method as claimed in claim 1 or 2, wherein, after the provision of the pocket implants, the second sub-layer is removed, and the conductive layer is applied.

4. A method as claimed in claim 3, wherein the transistor is applied to the surface of the semiconductor body as a p-channel transistor, and a silicon germanium alloy is applied as the first sub-layer.

5. A method as claimed in claim 4, wherein silicon oxide is applied as the dielectric layer, and silicon is applied as the second sub-layer.

6. A method as claimed in claim 2, wherein, after the removal of the patterned layer, an additional layer is applied, which additional layer provides the side walls of the recess aligned to the source zone and the drain zone.

7. A method as claimed in claim 6, wherein silicon is applied as the additional layer and the second sub-layer, and a silicon germanium alloy is applied as the first sub-layer.

8. A method as claimed in claim 6 or 7, wherein the double-layer and the additional layer are removed after the provision of the pocket implants, and the gate dielectric is provided at the surface of the semiconductor body, after which the conductive layer is applied.

9. A method as claimed in any one of the preceding claims, wherein the impurities are introduced into the semiconductor body by means of ion implantation.

10. A method as claimed in claim 9, wherein the ion implantation is carried out substantially perpendicularly to the surface of the semiconductor body.

11. A method as claimed in any one of the preceding claims, wherein the transistor is applied to the surface of the semiconductor body as an n-channel transistor having pocket implants formed by the introduction of boron or indium.

12. A method as claimed in any one of claims 1 to 10, wherein the transistor is applied to the surface of the semiconductor body as a p-channel transistor having pocket implants formed by the introduction of antimony or arsenic.

13. A method as claimed in any one of the preceding claims, wherein a layer comprising a metal is applied as the conductive layer.

10

17.02.2000

PH-NL000081 EP-P

14. A method as claimed in claim 13, wherein the layer comprising the metal is appl.ed as a double-layer consisting of a layer comprising the metal on top of a layer actir g as adhesion and/or barrier layer.

15. A method as claimed in claim 13 or 14, wherein the metal is selected from the group comprising aluminum, tungsten, copper, and molybdenum.